Molecular phylogenetics reveals a complex history underlying cryptic diversity in the Bush Squeaker Frog (*Arthroleptis wahlbergii*) in southern Africa

KA Tolley, W Conradie, J Harvey, J Measey & DC Blackburn

To cite this article: KA Tolley, W Conradie, J Harvey, J Measey & DC Blackburn (2018): Molecular phylogenetics reveals a complex history underlying cryptic diversity in the Bush Squeaker Frog (*Arthroleptis wahlbergii*) in southern Africa, African Zoology

To link to this article: [https://doi.org/10.1080/15627020.2018.1517608](https://doi.org/10.1080/15627020.2018.1517608)
Molecular phylogenetics reveals a complex history underlying cryptic diversity in the Bush Squeaker Frog (*Arthroleptis wahlbergii*) in southern Africa

KA Tolley<sup>1,2</sup>*, W Conradie<sup>3,4</sup>, J Harvey<sup>5</sup>, J Measey<sup>6</sup>, DC Blackburn<sup>7</sup>

1 South African National Biodiversity Institute, Kirstenbosch Research Centre, Claremont, Cape Town, South Africa
2 Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Auckland Park, Johannesburg South Africa
3 Port Elizabeth Museum, Humewood, South Africa
4 School of Natural Resource Management, Nelson Mandela University (George Campus), George, South Africa
5 Devonshire Avenue, Howick, South Africa
6 Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa
7 Florida Museum of Natural History, University of Florida, Gainesville, Florida, United States

*Corresponding author, e-mail: K.Tolley@sanbi.org.za

Throughout the Miocene, the African landscape underwent broad climatic shifts that profoundly influenced the distribution of fauna and flora. Since the late Miocene, these shifts have created a landscape in southern Africa that is strongly characterised by savanna and arid environments. Forests persist in small fragments, primarily in mountainous or heterogeneous landscapes. *Arthroleptis wahlbergii* is a small frog endemic to eastern South Africa that has presumed low dispersal ability. Because of its preference for forests, the dynamics of forests since the late Miocene in this region might have promoted diversification within *A. wahlbergii*. To investigate whether habitat fragmentation might have driven divergences among populations, we carried out species distribution modelling and population level phylogenetic analyses using two genetic loci (16S, mitochondrial; RAG-1, nuclear) sequenced for 48 individuals from 14 forests across the c. 500 km range of this species. There is substantial population-level structuring within *A. wahlbergii*, however the structure does not relate to forest types or catchments. We instead propose that the structure is a result of dynamic and idiosyncratic changes in forest connectivity over the Pleistocene. We identified two geographically circumscribed clades, the northern of which corresponds to true *A. wahlbergii*. The southern clade corresponds to populations from which *Arthroleptis wageri* FitzSimons, 1930 was described. This has long been considered a synonym of *A. wahlbergii*, but our molecular phylogenetic and distribution modelling supports recognising *A. wageri* as a distinct species.

Introduction

South Africa is home to three biodiversity hotspots that are characterised by high species richness for multiple floral and faunal groups (Mittermeier et al. 2004, 2011). The elevated species richness is undoubtedly the result of a complex climatic and geological history that provided opportunities for diversification and speciation (Cowling et al. 2009). In general, the southern African region has undergone long-term environmental changes that caused major biome reorganisation since the mid-Miocene (Zachos et al. 2001; Lawes et al. 2007; Kissling et al. 2012.), with a shift from mesic forest-dominated landscapes to more arid landscapes including savannah, grassland, and both Karoo and Mediterranean heathlands (Mucina and Rutherford 2006). This resulted in the contraction and fragmentation of forest into isolated patches that are maintained primarily by coastal or orographic rainfall. At present, the wet slopes of the Drakensberg escarpment and the Cape Fold Mountains support many of the forest patches in southern Africa, with the remainder scattered along the subtropical and tropical eastern seaboard (Mucina and Geldenhuys 2006). Although the general trend was for a reduction in forest, climatic oscillations with warmer, wetter periods would have allowed for temporary reconnections of forest patches, resulting in a complex history (Mucina and Geldenhuys 2006).

With the gradual fragmentation of forest habitat and replacement by more arid biomes, these novel landscapes might have been formidable barriers to movement of forest specialist species. Indeed, forest dependent species that do not adapt to the novel environment are typically found confined within forest refugia, leading to allopatric diversification (e.g. chameleons: Tolley et al. 2008, 2011). The simplest model of allopatric diversification relies on forest
fragmentation and long-term isolation in the relict patches. However, this simple model might not sufficiently explain the complex history of diversification of forest taxa. Species distribution models (SDMs) show that forest patches in South Africa have been spatially dynamic since the Last Glacial Maximum (LGM; 21 000 ybp) with numerous reconnections and contractions (Eeley et al. 1999). In some areas, forests might have recently (e.g. since the Holocene, c. 11 000 ybp) established in what was formerly grassland (West et al. 2000).

In South Africa, the extant relict forest patches are classified into four principle types based on floristic composition (Mucina and Geldenhuys 2006). Afrotemperate/Afromontane forest typically occurs at high elevation (1500–1900 m; scattered along a belt in the Drakensberg Mountains) and is floristically depauperate (Lawes et al. 2007). Mistbelt forest is found at lower elevation (850–1 600 m) and is floristically diverse, mostly with Afrotemperate elements (Lawes et al. 2007). Scarp forest occurs at low elevation (50–600 m), has exceptionally high species richness and endemism, and acted as a refuge during cooler climatic periods (Lawes et al. 2000). This forest type appears to be an amalgamation of Afrotemperate and coastal forests, although it is more similar to Coastal than to Afrotemperate/Mistbelt (Lawes et al. 2000; Mucina and Geldenhuys 2006; Lawes et al. 2007). Coastal forests, found at the lowest elevation (<350 m), are embedded within the larger Indian Ocean Coastal Belt biome that runs along the eastern seaboard. These forests are a recent southward expansion of Mozambique coastal forests (Lawes et al. 2007). They are more diverse in the north, owing to the occurrence of both tropical and subtropical species, with species richness declining at higher latitudes as tropical species reach their southern limits (Mucina et al. 2006). These small relict patches of each type were once presumably connected into larger and more contiguous forests.

There are a number of forest-associated vertebrate taxa in southern Africa, several of which are frog species that rely on stable humid environments. Arthroleptis is a genus of small-bodied (<30 mm) African frogs in the family Arthroleptidae, which are typically found in leaf litter on the forest floor or on forest edges (Poynton and Broadley 1985; Harper et al. 2010; du Preez and Carruthers 2018). They are direct developers with no free-swimming tadpole stage, laying small clutches (~10–50) of large-sized eggs (~2.5 mm) in leaf litter from which fully metamorphosed froglets hatch (Schweiger et al. 2017). Given their small size and requirement for a moist forest environment to maintain their egg clutches, most species are narrow range endemics often confined to individual mountains (e.g. Blackburn 2009; Harper et al. 2010). Conversely, some species are widespread with distributions that span savannah and montane habitats (e.g. Blackburn and Measey 2009). There are 47 Arthroleptis species in sub-Saharan Africa (Frost 2018), with only one species endemic to South Africa, Arthroleptis wahlbergii Smith, 1849, currently defining the southern range of this genus. This species is primarily distributed in highly fragmented Mistbelt, Scarp and coastal forest patches in eastern South Africa (Figure 1). It is occasionally found in adjacent thicket and grassland (Channing 2004) suggesting that it has some ability to cope with non-forest habitats, and this tolerance might facilitate gene flow between forests. SDMs for this species suggest that its distribution shifted substantially northwards during the LGM (21 000 ybp) and then moved southwards again during the Holocene (6 000 ybp; Schreiner et al. 2013). This generally reflects the forest SDMs that show a near complete loss of forest across much of the region during the LGM, but only re-established in patches during the Holocene (Eeley et al. 1999).

Recent molecular phylogenetic analyses revealed cryptic species within several lineages of Arthroleptis in which species are often highly similar to one another morphologically (Blackburn 2008; Blackburn et al. 2010). Indeed, given that subspecies have been previously recognised for A. wahlbergii (Loveridge 1954), it would not be surprising if some genetically divergent populations warrant recognition as separate species. For example, populations at the southern extent of the distribution were previously recognised as a separate species, Arthroleptis wageri (FitzSimons 1930), though it was later synonymised with A. wahlbergii (Poynton 1964). There are no previous assessments of population genetic structure and drivers of diversification among populations of A. wahlbergii. If some populations of A. wahlbergii were isolated in forest patches, there is the potential for allopatric diversification and cryptic speciation (see Struck et al. 2018). However, the positioning of forests within river catchments might be important in population-level structuring, as was found for the congener A. xenodactyloides from Kenya (Measey et al. 2007), as well as other amphibians and reptiles from East Africa (Measey and Tolley 2011; Lawson 2013). Therefore, forests that are within the same catchment potentially have higher connectivity and gene flow, and we might expect divergences corresponding to catchments.

Given the dynamics of forests in southern Africa and the presumed low dispersal of A. wahlbergii, we tested competing scenarios of diversification among populations. Initially, forest would have been fragmented by more arid biomes. If the fragmentation was relatively systematic across the landscape, neighbouring forests should harbour closely related lineages, and a pattern of isolation by distance would be evident (model a; Figure 2). Alternatively, if relationships are maintained through connectivity within catchments, we would expect forests within the same catchment to be closely related regardless of distance (model b; Figure 2). However, if the progress of vicariance was idiosyncratic with forests dissected in unpredictable ways, then nearest neighbours might not be closest relatives and isolation by distance would not be evident (model c; Figure 2). A final scenario is that populations in distinct forest types were isolated from each other, reducing gene flow across these different forests (model d; Figure 2). If so, closely related A. wahlbergii populations would occur only within a specific forest type with large differences between forest types.

These scenarios lead to different expectations for the patterns of phylogenetic relationships within A. wahlbergii across the current distribution of forests in eastern South Africa (Figure 2a–d). Furthermore, we hypothesised that
widely separated and isolated populations might have diverged in allopatry such that they are now genetically divergent, but morphologically similar species. We tested our predictions by constructing a molecular phylogeny of *Arthroleptis wahlbergii* using two genetic loci, and samples from multiple forest fragments from different forest types across its distribution. Lack of calibration points did not allow a dated phylogeny to be constructed, therefore we provide coarse estimates of divergence times using pairwise sequence divergences. We constructed new SDMs based on clades revealed by molecular phylogenetic analyses that aid in interpreting the drivers of divergence among populations.

**Materials and methods**

Field surveys for *A. wahlbergii* were carried out in 2011–2013 in 14 forest patches throughout KwaZulu-Natal.
and Eastern Cape Provinces, South Africa (Figure 1; Table 1). Forty-eight individual frogs were caught either by hand or pitfall traps set along drift fences. Thigh muscle or toe clips were taken for DNA extraction and stored in 99% ethanol (Venter and Conradie 2015).

To understand the evolutionary history of *A. wahlbergii*, a phylogenetic analysis using one mitochondrial (16S) and one nuclear (RAG-1) gene was carried out for a subset of 30 individuals of *A. wahlbergii* (Figure 1). Where possible, multiple individuals from each forest patch were included. DNA extraction, PCR amplification, and cycle sequencing were carried out following standard procedures using the following primers for 16S: L2510 and H3080 (Palumbi 2002), and for RAG-1: F118 and R1067 (Matthee et al. 2004). Standard PCR and sequencing were followed with annealing temperatures at 50–55 °C for 16S and 55–58 °C for RAG-1. The dataset included five additional *Arthroleptis* species that are closely related to *A. wahlbergii* and for which both loci were available via GenBank: *A. adelphus, A. reichei, A. sylvaticus, A. tanneri*, and *A. variabilis*, with *A. francei* as the outgroup taxon (Blackburn 2008). Both loci were edited for ambiguities and aligned in Geneious v8 using the MUSCLE alignment plug-in (Kearse et al. 2012). All new sequences were deposited in GenBank (Table 1).

Bayesian inference was used to investigate optimal tree space using MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) on the dataset of 1 464 characters, partitioned by gene (16S, 430 bp; RAG-1, 1 034 bp), although 8 bases were excluded for 16S, because of ambiguous alignment.

Figure 2: Four scenarios of vicariance for *Arthroleptis wahlbergii*. Forest types are denoted by colour: Mistbelt (green), Scarp (blue) and Coastal (orange). Model a) systematic fragmentation; model b) associations via watershed; model c) idiosyncratic fragmentation; model d) gene flow within forest types, with no gene flow among forest types. Stages over time are represented from left to right, and the final stage has identical spatial positioning of forest patches across scenarios. Gene flow is presumed to be reduced across unsuitable habitat (darker shading). Dotted lines on right panel indicate hypothetical closest relationships or groupings.
To investigate which evolutionary model best fit the data, jModeltest was used (Posada 2008); the AIC test indicated the same model for both genes (GTR + G; a = 0.25) was most appropriate. This differed from the BIC test, which showed that a simpler model was more appropriate for the 16S marker (HKY+G). Therefore, MrBayes was run specifying two rate categories for 16S (favouring the simpler model) and six rate categories for RAG-1, with uniform priors for the gamma distribution for each of the partitions. To ensure the results were robust, the MCMC was run twice in

Table 1: List of samples used in the genetic analyses, including sampling locality, forest type, clade membership and GenBank accession numbers for 16S and RAG-I for *Arthroleptis wahlbergii* and outgroup taxa. Also indicated is the use of each sample in the phylogeny, 16S network or RAG-1 network. Y = sample included, na = not included.

<table>
<thead>
<tr>
<th>Field number</th>
<th>Species</th>
<th>Locality</th>
<th>Forest type</th>
<th>Clade</th>
<th>16S</th>
<th>RAG-I</th>
<th>Phylogeny</th>
<th>16S network</th>
<th>RAG-I network</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB002</td>
<td><em>A. wahlbergii</em></td>
<td>Blythdale Beach</td>
<td>Coastal</td>
<td>D</td>
<td>MH729272</td>
<td>MH729241</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>BB003</td>
<td><em>A. wahlbergii</em></td>
<td>Blythdale Beach</td>
<td>Coastal</td>
<td>D</td>
<td>MH729273</td>
<td>MH729242</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>EN001</td>
<td><em>A. wahlbergii</em></td>
<td>Ntumeni</td>
<td>Scarp</td>
<td>G</td>
<td>MH729295</td>
<td>MH729254</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>EN002</td>
<td><em>A. wahlbergii</em></td>
<td>Ntumeni</td>
<td>Scarp</td>
<td>G</td>
<td>na</td>
<td>MH729255</td>
<td>na</td>
<td>na</td>
<td>Y</td>
</tr>
<tr>
<td>HB194</td>
<td><em>A. wahlbergii</em></td>
<td>New Hanover Mistbelt</td>
<td>F</td>
<td>MH729280</td>
<td>MH729256</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>HB195</td>
<td><em>A. wahlbergii</em></td>
<td>New Hanover Mistbelt</td>
<td>F</td>
<td>MH729281</td>
<td>MH729257</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>HB198</td>
<td><em>A. wahlbergii</em></td>
<td>Dlinza</td>
<td>Scarp</td>
<td>D</td>
<td>MH729274</td>
<td>MH729258</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>HB199</td>
<td><em>A. wahlbergii</em></td>
<td>Dlinza</td>
<td>Scarp</td>
<td>D</td>
<td>MH729275</td>
<td>MH729259</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>HB200</td>
<td><em>A. wahlbergii</em></td>
<td>Dlinza</td>
<td>Scarp</td>
<td>D</td>
<td>MH729276</td>
<td>MH729260</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>HB201</td>
<td><em>A. wahlbergii</em></td>
<td>Nkandla</td>
<td>Scarp</td>
<td>E</td>
<td>MH729293</td>
<td>MH729261</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>HB203</td>
<td><em>A. wahlbergii</em></td>
<td>Isipingo</td>
<td>Coastal</td>
<td>F</td>
<td>MH729277</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>HB204</td>
<td><em>A. wahlbergii</em></td>
<td>Ngome</td>
<td>Mistbelt</td>
<td>B</td>
<td>MH729282</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>HB228</td>
<td><em>A. wahlbergii</em></td>
<td>Treasure Beach</td>
<td>Coastal</td>
<td>F</td>
<td>MH729306</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>HW001</td>
<td><em>A. wahlbergii</em></td>
<td>Hluhluwe</td>
<td>Scarp</td>
<td>C</td>
<td>MH729305</td>
<td>MH729263</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAT13001</td>
<td><em>A. wahlbergii</em></td>
<td>oNgoya</td>
<td>Scarp</td>
<td>D</td>
<td>MH729288</td>
<td>MH729264</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAT13002</td>
<td><em>A. wahlbergii</em></td>
<td>oNgoya</td>
<td>Scarp</td>
<td>D</td>
<td>MH729289</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>JHAT13003</td>
<td><em>A. wahlbergii</em></td>
<td>Qudeni</td>
<td>Mistbelt</td>
<td>E</td>
<td>MH729307</td>
<td>MH729266</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAT13004</td>
<td><em>A. wahlbergii</em></td>
<td>Qudeni</td>
<td>Mistbelt</td>
<td>E</td>
<td>MH729308</td>
<td>MH729266</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAT13005</td>
<td><em>A. wahlbergii</em></td>
<td>Qudeni</td>
<td>Mistbelt</td>
<td>E</td>
<td>MH729309</td>
<td>MH729266</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAT13006</td>
<td><em>A. wahlbergii</em></td>
<td>Qudeni</td>
<td>Mistbelt</td>
<td>E</td>
<td>MH729310</td>
<td>MH729267</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAT13009</td>
<td><em>A. wahlbergii</em></td>
<td>Nsezi</td>
<td>Coastal</td>
<td>C</td>
<td>MH729301</td>
<td>MH729252</td>
<td>Y</td>
<td>na</td>
<td>Y</td>
</tr>
<tr>
<td>JHAW1</td>
<td><em>A. wahlbergii</em></td>
<td>Isipingo</td>
<td>Coastal</td>
<td>F</td>
<td>MH729278</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAW1a</td>
<td><em>A. wahlbergii</em></td>
<td>Isipingo</td>
<td>Coastal</td>
<td>C</td>
<td>MH729302</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>JHAW2</td>
<td><em>A. wahlbergii</em></td>
<td>Isipingo</td>
<td>Coastal</td>
<td>F</td>
<td>MH729279</td>
<td>MH729253</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JHAW2b</td>
<td><em>A. wahlbergii</em></td>
<td>Isipingo</td>
<td>Coastal</td>
<td>C</td>
<td>MH729300</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>JHAW3</td>
<td><em>A. wahlbergii</em></td>
<td>Nsezi</td>
<td>Coastal</td>
<td>C</td>
<td>MH729303</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>JHAW4</td>
<td><em>A. wahlbergii</em></td>
<td>Nsezi</td>
<td>Coastal</td>
<td>C</td>
<td>MH729304</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>JHAW5</td>
<td><em>A. wahlbergii</em></td>
<td>Ntumeni</td>
<td>Scarp</td>
<td>G</td>
<td>MH729296</td>
<td>na</td>
<td>Y</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>JHAW6</td>
<td><em>A. wahlbergii</em></td>
<td>Ntumeni</td>
<td>Scarp</td>
<td>G</td>
<td>MH729297</td>
<td>na</td>
<td>Y</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>JHAW7</td>
<td><em>A. wahlbergii</em></td>
<td>oNgoya</td>
<td>Scarp</td>
<td>D</td>
<td>MH729292</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>NK001</td>
<td><em>A. wahlbergii</em></td>
<td>Nkandla</td>
<td>Scarp</td>
<td>E</td>
<td>MH729294</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>NK002</td>
<td><em>A. wahlbergii</em></td>
<td>Nkandla</td>
<td>Scarp</td>
<td>E</td>
<td>MH729295</td>
<td>na</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
<tr>
<td>PEM A10001</td>
<td><em>A. wageri</em></td>
<td>Hluleka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729308</td>
<td>MH729243</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10500</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729309</td>
<td>MH729244</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10548</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729310</td>
<td>MH729245</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10549</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729311</td>
<td>MH729246</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10551</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729312</td>
<td>MH729247</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10552</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729313</td>
<td>MH729248</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10553</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729314</td>
<td>MH729249</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A10554</td>
<td><em>A. wageri</em></td>
<td>Silaka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729315</td>
<td>MH729250</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PEM A9967</td>
<td><em>A. wageri</em></td>
<td>Hluleka</td>
<td>Scarp</td>
<td>A</td>
<td>MH729307</td>
<td>MH729251</td>
<td>na</td>
<td>Y</td>
<td>na</td>
</tr>
</tbody>
</table>

Outgroup:

- **MCZ A136732**  A. adelphus  Cameroon, Mangome  na  na  FJ151081  MH744348  Y  na  na
- **MCZ A137038**  A. francis  Malawi, Mt. Mulanje  na  na  FJ151100  MH744349  Y  na  na
- **MCZ A138365**  A. reichei  Tanzania, Udzungwa  na  na  FJ151151  MH744350  Y  na  na
- **MVZ 234674**  A. sylvaticus  Cameroon, Nguti  na  na  FJ15106  MH744351  Y  na  na
- **CAS 168823**  A. tanneri  Tanzania, W. Usambara  na  na  FJ151056  MH744352  Y  na  na
- **CAS199161/ZFMK68794**  A. variabilis  Cameroon  na  na  EU350212  AY571642  Y  na  na
parallel for 20 million generations (four chains in each run), with trees sampled every 1,000 generations. The effective sample sizes (ESS) of all parameters were checked in Tracer v1.4 (Rambout and Drummond 2007) to ensure they met a threshold of ESS >200 after a 10% burn-in was discarded (2 million generations, 2,000 trees). A 50% majority rule tree was constructed and nodes with ≥0.95 posterior probability considered supported.

In addition to the Bayesian analysis, a maximum likelihood (ML) search was run using RAxML HPC v7.2.8 (Stamatakis 2006) on the CIPRES Science Gateway (www.phylo.org/sub_sections/portal/) for the combined dataset. The datasets were partitioned as in the Bayesian analysis and the default GTR+I+G model was used for both genes with rapid bootstrapping halted automatically (Stamatakis et al. 2008). This analysis was run three times to ensure that independent ML searches produced the same topologies. We considered nodes with a bootstrap value of ≥70%, as supported in this analysis.

Pairwise sequence divergence values (uncorrected net p-distances, pairwise deletion of missing data) were estimated between *A. wahlbergii* clades for the 16S gene using MEGA v7.0.14 (Kumar et al. 2016). Coarse estimates of the time periods associated with divergence between clades were made by applying a mutation rate of 1% per million years per lineage to the divergence estimates (following Jongsma et al. 2018). These estimates provide a rough estimation of the epoch during which divergences occurred (e.g., Holocene, Pleistocene, Pliocene, Miocene).

A barcoding approach to species delimitation was used to compare inter- and intraspecific sequence divergences using SpeciesIdentifier v1.8 (Meier et al. 2006). A frequency distribution of pairwise divergence values was generated for 163 individuals of *Arthroleptis* from 26 species (downloaded from GenBank) for the 16S gene. The dataset (558 base pairs) for this diverse set of species had regions of poor alignment, consequently 107 base pairs were selected for exclusion using relaxed settings (Castresana 2000; Talavera and Castresana 2007) resulting in the inclusion of maximum 451 base pairs per individual. The inter- and intraspecific sequence divergences were binned separately to generate inter- and intraspecific frequency distributions. The divergences among populations of *A. wahlbergii* were then compared to those two distributions to determine whether they fell in the range of inter- or intraspecific values.

A median-joining network was constructed for each gene separately to examine reticulated relationships and allele sharing between localities using Network v5 (Bandelt et al. 1999). The 16S dataset was truncated to 357 base pairs and the RAG-1 dataset was truncated to 505 base pairs to remove missing data (Table 1). Ambiguity codes were used where two alleles were present for RAG-1 (two bases present at a single base site). A number of additional individuals were sequenced for 16S and these were included in the network, but not the phylogeny (Table 1). Isolation by distance (IBD) was examined by estimating genetic and geographic distances between all individuals of *A. wahlbergii* using Alles in Space (Miller 2005). Genetic distances were generated from the 16S sequences and geographic distances were estimated using the collection locality coordinates. We then used a Mantel test to examine whether there was significant IBD.

The degree of differentiation among the forests was also assessed using a spatial analysis of molecular variance (AMOVA v2; Dupanloup et al. 2002) for the 16S gene. AMOVA uses the geographical coordinates and corresponding sequence data to assign sampling sites a posteriori to groups that presumably represent historically cohesive populations. The procedure uses a simulated annealing process to maximise the proportion of total genetic variation between groups of sample sites using traditional F statistics. F_{CT} is the proportion of total genetic variance attributable to the differences between groups of sample sites; F_{ST} represents the variation between sample sites within groups whereas F_{IS} is the genetic variation between sample sites relative to the total sample (Excoffier et al. 2005; Dupanloup et al. 2002). AMOVA was run for all possible group numbers (K) to determine the optimal value for F_{CT}.

We also grouped forests *a priori* according to the primary and secondary catchments where they occur. Both assignments were run in a hierarchical analysis of molecular variance (AMOVA) to estimate F_{CT} values constrained by these two catchment levels. These values were compared with the AMOVA values to examine whether catchments maintain relationships between the populations.

Species distribution models

Our approach for species distribution models followed Schreiner et al. (2013), although differed somewhat as a result of being informed by our phylogenetic analysis (see Results and Discussion). We divided occurrence records into those corresponding with *A. wahlbergii sensu stricto* and the southernmost populations corresponding to *A. wageri sensu FitzSimons* (1930). We used occurrence records from our field surveys (see above) and added data from the South African Frog Atlas Project (Minter et al. 2004) that are accurate to the nearest second or higher. Distribution records were projected using ArcMap GIS v10.3 (ESRI 2014) to identify dubious records, such as those in the sea or outside known ranges, which were excluded from subsequent analyses. This provided 278 records for ‘*A. wahlbergii*’ and 25 for ‘*A. wageri*.’

Environmental predictors were compiled from a set of 19 bioclimatic variables at a spatial resolution of 2.5 arc minutes available from WorldClim (Hijmans et al. 2005). Intercorrelation structure among predictors throughout the study area was assessed by computing pairwise squared Spearman’s rank correlation coefficients, and in cases where R^2 exceeded 0.75, only the putatively biologically most important variables were retained. Accordingly, ten variables were retained: annual mean temperature (BIO1), mean diurnal range (mean of monthly [maximum temperature − minimum temperature; BIO2]), isothermality (mean diurnal range/temperature annual range; BIO3), temperature annual range (BIO7), temperature of wettest quarter (BIO8), mean temperature of coldest quarter (BIO11), annual precipitation (BIO12), precipitation seasonality (BIO15), precipitation of warmest quarter (BIO18) and precipitation of coldest quarter (BIO19).

Species distribution model projections were created at a 2.5 arc minute resolution. To reconstruct past distributions,
we projected SDMs onto paleo-climate for the Last Glacial Maximum (LGM = 21 kya), using the Community Climate System Model (CCSM: Otto-Bliesner et al. 2006) general circulation models (GCMs). Unlike Schreiner et al. (2013), we chose to run only CCSM models for LGM scenarios, because these had been found to perform better. Paleo-climate data were set at a spatial resolution of 2.5 arc minutes following the delta method (Peterson and Nyari 2008). Current distributions were based on SDM projections into mean climate from 1950 to 2000. Species distribution modelling was done using the machine learning algorithm Maxent v3.3.3k (Phillips and Dudik 2008), based on the ten bioclimatic variables. Maxent has been successfully used to model past and current distributions of endemic southern African frog species (e.g., Schreiner et al. 2013). Distribution records were divided into 80% used to train the models and 20% for model evaluation using the area under the receiver operating characteristic curve: AUC (Swets 1988). The final model was averaged over 100 replicates, using the bootstrap sampling technique. A circular buffer of 200 km around each locality point was selected as the training area following Measey et al. (2012). When projecting the ensemble across space and time, non-analogous climatic conditions that exceeded the training range of the models were quantified wherein the numbers of these predictors are highlighted using Multivariate Environmental Similarity Surfaces (MESS: Elith et al. 2011). We used the “Minimum training presence threshold” referring to the lowest generated probability estimates of the training data (Pearson et al. 2007) with suitable predictions predicted above the threshold and unsuitable predictions below the threshold.

Results

*Arthroleptis wahlbergii* was recovered as monophyletic with respect to other species of *Arthroleptis*, with high support in both analyses (Figure 3). Within *A. wahlbergii*, there was a well-supported split between individuals from the southern forests (clade A) and those from the northern forests (clades B–G). In addition, several clades within the northern clade were well supported, though nodes within the northern clade lacked strong support. The monophyly of the northern clade received modest support (66% ML, 0.91 Bayesian pp).

As with the phylogenetic analyses, the median-joining network for 16S shows numerous mutations separating clade A (southern) from clades B–G (northern; Figure 4a). The haplotypes within clades B–G are more closely related, and there are multiple reticulations. However, there are essentially no shared haplotypes between forest patches, with the exception of Dlinza and oNgoya. For RAG-1, there is substantial allele sharing between localities in clades B–G, and one individual from clade F shares an allele with clade A (Figure 4b).

Uncorrected net p-distances for 16S reveal the southern clade A to be c. 4–6% divergent from the northern clades B–G (Table 2). Applying the constant mutation rate of 1% per million years per lineage, the southern (A) and northern (B–G) clades would have diverged roughly 2–3 million ybp in the late Pliocene or early Pleistocene. Divergences between clades within the north are roughly half the divergence between the northern and southern clades, suggesting that these clades diverged in the mid-Pleistocene (c. 1–1.5 million ybp). The barcoding analysis indicates that the divergence between the northern and southern clades falls firmly within the range of interspecific divergences typical within *Arthroleptis* (Figure 5). Clades B–G (within the larger northern clade) all show divergences that correspond with values that encompass inter- and intraspecific divergences.

There was significant isolation by distance (IBD) within *A. wahlbergii* (r = 0.32, p < 0.001). However, if individuals from the southern clade A were removed from the analysis, there was no IBD (r = −0.35, ns) within the northern clade (B–G), indicating that the strong divergence between northern and southern forests drives the IBD pattern. The SAMOVA showed the variance between forests is maximised where the value of FCT reaches a plateau at K = 6 (Figure 6). Assuming the most parsimonious group structure is six, this supports the phylogenetic analysis, with the only difference being that clades E and F are in the same SAMOVA grouping. The assignment of forests to catchments showed much lower FCT values (Figure 6) for primary and secondary catchments than the SAMOVA groupings.

Species distribution models

We found that both ‘species’ performed well in modelled current distributions with high Area Under the Curve (AUC) values for both training and test models: *A. wageri sensu* FitzSimons (1930) AUC scores = 0.959 (AUC\textsubscript{training}) 0.957 (AUC\textsubscript{test}); *A. wahlbergii* AUC scores = 0.909 (AUC\textsubscript{training}) 0.905 (AUC\textsubscript{test}). Both species had similar contributions from complementary BIOCLIM climatic variables (Table 3). In accordance with Schreiner et al. (2013), our models predicted the same reduced potential distribution from 21 kya to present. However, the use of *A. wahlbergii sensu stricto* data provided a far more patchy distribution away from the current coastal range, whereas the models for *A. wageri sensu* FitzSimons (1930) suggest the distribution in suitable climate is relatively limited and has not changed substantially since the LGM (Figure 7).

Discussion

Phylogenetic patterns

Phylogenetic analyses support a strong divergence within *A. wahlbergii* between the populations in southern forests (clade A) and those in northern forests (clades B–G). Because the barcoding analysis supports this divergence as comparable with species-level differences observed between other species of *Arthroleptis*, we propose that these two reciprocally monophyletic clades represent separate species. However, the northern and southern clades share at least one nuclear RAG-1 allele. Although the nuclear allele sharing is not widespread, it does suggest some degree of historical connectivity between these regions. Within clade A, the two southern coastal forest sample sites (Silaka and Hluleka) separated by c. 25 km show no phylogenetic structure, suggesting recent or current connectivity in this area. Indeed, there are multiple additional coastal forest patches embedded in the intervening landscape composed of Indian Ocean Coastal
Figure 3: Maximum likelihood topology for *Arthroleptis wahlbergii*, and other representatives of the genus. Nodes supported by both maximum likelihood (≥ 70% bootstrap) and Bayesian (≥ 0.95 posterior probability) analyses are denoted with black circles, those supported by maximum likelihood only are denoted with white circles, and those supported by Bayesian only are denoted with grey circles. Node with an asterisk is supported by 66% bootstrap and 0.91 posterior probability. Vertical bars are colour coded to indicate forest types for each sampling locality: Mistbelt (green), Scarp (blue) and Coastal (orange). Primary catchment names are given for each clade, and the geographic localities for catchments are indicated in the inset map.
Belt (IOCB). Furthermore, the IOCB is itself characterised by a complex savannah/forest vegetation matrix, and is usually considered as an additional forest type (Mucina et al. 2006; Lawes et al. 2007). Given that the IOCB and coastal forests are essentially a mosaic of different kinds of forest, there should be ample pathways for dispersal within the ‘forest-like’ vegetation. Moreover, this species could even occur within the non-forest IOCB vegetation (the savanna component) which would facilitate gene flow. In contrast, the 12 northern forests show substantial phylogenetic structure with multiple well-supported clades (B–G). However, several of these clades contain samples from multiple forests that are currently isolated. This suggests either historical connections between such forests, or that these frogs can disperse across the less suitable savanna landscape (e.g. C: Nsezi and Hluhluwe, D: oNgoya, Dlinza and Blythdale, E: Qudeni and Nkandla; see Measey et al. 2007).

**Taxonomic issues**

Our genetic results suggest that the populations in the southern and northern forests represent candidate species. The southern clade (A: Silaka and Hluleka forests) is well supported as distinct and is differentiated from the northern clade in the barcoding analysis. However, the northern clade is only moderately supported as monophyletic (66% ML bootstrap, 0.91 Bayesian posterior probability). In addition, the common RAG-1 allele in the southern forests is shared with one northern forest (New Hanover). This is unlikely the result of contemporary gene flow, given that 16S haplotypes are highly divergent between the southern and northern clades, suggesting instead that there has been retention of ancestral alleles for RAG-1 in these forests. Currently the barcoding results, the mitochondrial divergence, the low frequency of allele sharing for the nuclear gene and the phylogeny (albeit with low support for monophyly of the northern clade) all lend support to recognizing two distinct species.
boundaries, as well as the phenotypic and advertisement of other populations will help to delimit the geographic and typical our decision to recognise (1930) from synonymy with A. wahlbergii. Based on our results, we elevate FitzSimons (1930) described A. dalenei from near Greytown from which was more than one species. Although we have no samples morphological differences did not warrant recognition of limb length, though Poynton (1964) argued that these provided morphological criteria for distinguishing FitzSimons (1930), Loveridge (1954), and Laurent (1961) because the reserve is within 5 km of the town centre. A. wageri Silaka Nature Reserve to be topotypic for, from Port St. Johns, and we consider our samples from A. wageri based on a small series of material described, its location (c. 45 km SSW of Qudeni forest) Arthroleptis dalenei from near Greytown from which was recognised as a distinct species. We acknowledge that our study did not include analyses of morphological or advertisement calls; however, the phylogenetic results strongly support our decision to recognise A. wageri. Additional examination of other populations will help to delimit the geographic boundaries, as well as the phenotypic and advertisement differences between A. wahlbergii and A. wageri, which is now the southernmost species of the genus.

**Historical Connectivity**

Species distribution model predictions suggest that forest patches were highly reduced in extent during the Last Glacial Maximum (LGM: c. 18 000 ybp; Eeley et al. 1999), in the same way that our SDM predict a more patchy distribution for A. wahlbergii at the LGM. It is likely that Ngome Forest remained essentially intact, although it most likely shifted downwards in elevation (and therefore eastwards) and would have been completely isolated from forests to the south (Eeley et al. 1999). Indeed, this matches our genetic and SDM results, whereby the northern clade B contains only individuals from Ngome Forest, which are expected to have been isolated from other populations at the LGM (Figure 7), although divergence values between Ngome and other northern forests are not notably larger. In addition, an east–west archipelago of forest patches probably persisted in the area that now includes Dinza, Nkandla and Qudeni (Eeley et al. 1999). It is possible that within that forest archipelago, gene flow occurred between Nkandla and Qudeni (both in clade E). These two disjunct forests also contain a single species of chameleon (Bradypodion nemorale) that shows population level divergence (Tilbury and Tolley 2009). As with A. wahlbergii and A. wageri, species of Bradypodion

![Figure 5: Frequency distribution of pairwise differences for the mitochondrial 16S gene between Arthroleptis individuals representing 26 species. Interspecific values are in light grey and intraspecific values in dark grey. The level of sequence divergence within and between southern (a) and northern (b–c) clades for Arthroleptis wahlbergii is denoted.](image)
have low dispersal ability outside their preferred habitat. The evidence from both chameleons and *Arthroleptis* points to a historical connection between Nkandla and Qudeni, despite these currently being disjunct and composed of different forest types. Furthermore, Dlinza Forest is strongly associated with oNgoya Forest for *A. wahlbergii*, as well as chameleons, with a single species (*B. caeruleogula*) shared across the two forest patches. The SDM models also indicate that oNgoya Forest might not have been intact during the LGM, but that instead is the result of expansion of forests in this region during the Holocene altithermal (c. 7 000 ybp; Eeley et al. 1999). Given that these two forests are the only ones that share *A. wahlbergii* haplotypes, the most parsimonious explanation is that there was a larger patch of forest during the Holocene altithermal (Dlinza + oNgoya) that has since fragmented into two parts. The forest SDM suggests that most other patches of Mistbelt and Scarp forests were at a tiny fraction of current extent, and coastal forest was not intact (Eeley et al. 1999).
Figure 7: Species distribution models for a) *Arthroleptis wageri* sensu FitzSimons (1930) and b) *Arthroleptis wahlbergii* corresponding to current climate scenarios (Current) and hind casting to conditions at the Last Glacial Maximum (LGM) using Community Climate System Model (CCSM: see Methods for details).
Certainly, coastal forest is thought to be a young forest type (Mucina and Geldenhuys 2006; Lawes et al. 2007), which fits the forest SDM model predictions at the LGM. Coastal forest would have developed during the Holocene altithermal (Eeley et al. 1999), expanding southward from Mozambique (Lawes et al. 2007). Populations of *A. wahlbergii* from coastal forests (Isipingo/Treasure Beach, Blythedale, Nsezi) are all more closely related to the nearest inland forest (either Scarp or Mistbelt) than to each other. Therefore, scarp forests might have acted as refugia for *A. wahlbergii*, which later colonised eastward into coastal forest during the Holocene altithermal. Similarly, SDM predictions from chameleons (B. melanocephalum) show that the coastal areas have low climatic stability and that chameleons were excluded from the coast during the LGM (da Silva and Tolley 2017).

For the northern clade, the inland forests might have acted as refugia that expanded into coastal forest as it became available. Habitat fragmentation into multiple refugia during glacial maxima, with subsequent re-expansion has been noted for a number of other amphibians (e.g. Wang et al. 2017, Zimkus et al. 2017). Indeed, palaeoclimatic species distribution models for *A. wahlbergii* show the potential distribution as reduced and fragmented during the Last Glacial Maximum (LGM; c. 21 000 ybp), compared with the current situation (Schreiner et al. 2013). The SDM suggests that interior regions, particularly in the north, would have provided suitable climatic space probably as multiple refugia. Interestingly, our SDM for *A. wahlbergii* suggests there could have been a refugial southern population near Mahlabathini during the LGM (Figure 7, southernmost suitable area in LGM panel). We included no samples from this area, but this would be an important population to investigate in future. Suitable climate probably became available again at Holocene altithermal (c. 6 000 ybp) along the coast, which fits with an inland refugial hypothesis.

Because our species-level taxon sampling of *Arthroleptis* was incomplete and there are no reliable calibration points from fossils within this genus, estimating divergence times was not possible. However, our coarse estimates point to the divergence between northern and southern clades around the Plio–Pleistocene transition, with divergences among the northern forests well within the Pleistocene. Therefore, changes in forest extent across this larger landscape associated with global cooling and aridification at the start of the Pleistocene (deMenocal 1995, 2004) might have driven diversification between north and south. The paleo-climatic models projected to the LGM, however, suggest that there might have been substantial changes in the extent of distribution, and we can infer that these changes would have been similarly dynamic throughout the Pleistocene. Interestingly, our palaeoclimatic model for *A. wageri sensu* FitzSimons (1930) differs substantially from that of *A. wahlbergii*, by suggesting that *A. wageri* did not experience fragmentation at the LGM. This contrasts with the continuous distribution predicted by Schreiner et al. (2013) when they used a combined dataset of populations representing both *A. wahlbergii* and *A. wageri*. Although the genetic divergence we observed between these two species is deep and likely driven by environmental changes prior to the LGM, the palaeoclimatic models demonstrate that the connections between north and south have not been persistent. We therefore suggest that habitat refugia, where they existed, would have been largely disconnected between north and south for much of the Pleistocene.

**Vicariance models**

Three of the four vicariance models proposed (Figure 2) are not supported by our results. Firstly, systematic fragmentation (model a) can be ruled out, because of the lack of isolation by distance, and because nearest neighbours are not necessarily within the same clade or SAMOVA group. For example, specimens from oNgoya and Nsezi are in different clades and SAMOVA groups, though they are just 25 km apart. In turn, each of their closest relatives are up to 70 km away. In a more extreme case, Dlinza and Ntumeni are only 6 km apart, but are not closest relatives. Regardless, in some cases nearest neighbours are close relatives (e.g. Qudeni and Nkandla, Isipingo and Treasure Beach), accordingly would be expected to have maintained historical connections. However, given that frogs from these forests do not share mitochondrial haplotypes (Table 1), it seems unlikely there is substantial gene flow although additional sampling could reveal some shared haplotypes.

Model b, in which connections are maintained within primary catchments, can also be ruled out. This model showed much lower $F_{CT}$ values than the SAMOVA groupings and had no correspondence with the clades recovered in phylogenetic analysis. Furthermore, Dlinza and Ntumeni (close neighbours, but not close relatives) are not only in the same primary catchment, but are also in the same tertiary catchment. While some clades contain forests in the same catchment (e.g. clade A, E, F; Figure 3), there is no overall consistent pattern suggesting that structure is not maintained through the catchments.

The reduced gene flow across forest types (model d) can be ruled out. Most clades contain sampling sites from multiple forest types, and although each clade usually has members from spatially adjoining forest types (e.g. Qudeni from Mistbelt and Nkandla from Scarp), this is not always the case (e.g. New Hanover from Mistbelt and Isipingo from Coastal).

Our results best support an idiosyncratic process of vicariance (model c), reflected by the intricate pattern of forest types falling across the phylogeny, as well as the lack of isolation by distance. The SAMOVA failed to group forests according to any predictable pattern (e.g. by geographic distance, catchment or forest type). The forest fragments in this region have a dynamic history with expansions, contractions and shifts in distribution.

**Ethical clearance** — Ethical clearance for this study was obtained from the permits office of the South African National Biodiversity Institute for funding lab work and field work. We thank Jan Venter (Eastern Cape Parks and Tourism Agency) for facilitating fieldwork along the Wild Coast, and Adrian Armstrong and Jessica da Silva for facilitating fieldwork.
and carrying out sampling in KwaZulu-Natal. We thank Shandre Dreyer and Tes Linevee for assistance in the laboratory and Mohlamatsane Mokhatla for assistance with the species distribution models.

**ORCID**

KA Tolley [https://orcid.org/0000-0002-7778-1963](https://orcid.org/0000-0002-7778-1963)
W Conradie [https://orcid.org/0000-0003-0805-9683](https://orcid.org/0000-0003-0805-9683)
J Measey [https://orcid.org/0000-0001-9939-7615](https://orcid.org/0000-0001-9939-7615)
DC Blackburn [https://orcid.org/0000-0002-1810-9886](https://orcid.org/0000-0002-1810-9886)

**References**


